

# An Efficient Parameterization of the Room Transfer Function

Prasanga Samarasinghe\*, *Student Member, IEEE*, Thushara Abhayapala, *Senior Member, IEEE*,  
Mark Poletti, *Senior Member, IEEE*, and Terence Betlehem, *Senior Member, IEEE*

**Abstract**—This paper proposes an efficient parameterization of the Room Transfer Function (RTF). Typically, the RTF rapidly varies with varying source and receiver positions, hence requires an impractical number of point to point measurements to characterize a given room. Therefore, we derive a novel RTF parameterization that is robust to both receiver and source variations with the following salient features: (i) The parameterization is given in terms of a modal expansion of 3D basis functions. (ii) The aforementioned modal expansion can be truncated at a finite number of modes given that the source and receiver locations are from two sizeable spatial regions, which are arbitrarily distributed. (iii) The parameter weights/coefficients are independent of the source/receiver positions. Therefore, a finite set of coefficients is shown to be capable of accurately calculating the RTF between any two arbitrary points from a pre-defined spatial region where the source(s) lie and a pre-defined spatial region where the receiver(s) lie. A practical method to measure the RTF coefficients is also provided, which only requires a single microphone unit and a single loudspeaker unit, given that the room characteristics remain stationary over time. The accuracy of the above parameterization is verified using appropriate simulation examples.

## I. INTRODUCTION

The room transfer function (RTF), demonstrates the collective effect of multipath propagation of sound between a source and a receiver within a given room enclosure. Accurate modeling of the RTF is useful in soundfield simulators as well as many other applications such as sound reproduction, soundfield equalization, echo cancellation, and speech dereverberation. These applications use appropriate RTF deconvolution methods to cancel the effects of room reflections (reverberation), and therefore, are highly dependent on the accuracy of the RTF model.

The theoretical solution to the RTF based on the Green's function [1] was derived assuming a strict rectangular room geometry. It can only be applied to highly idealised cases with reasonable effort. The rooms with which we are concerned in our daily life however are more or less irregular in shape and the formulation of irregular boundary conditions will require extensive numerical calculations. For this reason, the immediate application of the classical model to practical problems in room acoustics is limited.

In practice, RTFs are usually estimated as FIR filters, or as parametric equations based on the geometrical properties of the room. In the FIR filter approach, the RTF is assumed to behave as a linear time-invariant system, and then modeled as either an all-zero, all-pole, pole-zero [2] or a common pole-zero [3] system. The coefficients of these models are estimated as variable parameters of the RTF, and since the

RTF is extremely sensitive to source and receiver variations, the coefficients too experience a similar sensitivity [3]. This problem not only requires repetitive parameter calculations with varying source/receiver positions, but also demands for adaptive inverse-filters with cumbersome processing algorithms during equalization [4], [5]. Furthermore, in practice, the time invariant aspect of room acoustics is far from reality [5], which remains as a fundamental weakness of the time-invariant filter model.

In contrast, the geometric room acoustics model, heavily relies on the room geometry and ray optic methods borrowed from computer graphics. The first geometric model for room reverberation was introduced by Allen and Berkley [6]. This work became the basis for many subsequent geometric models and is based on the notion that reverberation can be represented as the effect of an infinite number of image sources that are created by reflecting the true acoustic source in room walls. A faster algorithm to evaluate the image source method for single source-multiple receiver applications was later introduced in [7] using the multipole expansion. Other common geometric models include ray tracing [8], beam tracing [9], acoustic radiosity [10], and Finite Difference Time Domain (FDTD) [11], [12] methods. Even though these techniques have certain similarities, their theoretical foundations are often unique for each method. For example, the ray tracing method assumes high operating frequencies while the FDTD method assumes a low-mid frequency bandwidth [11]<sup>1</sup>. Therefore, their applicability to a general room is quite limited. More generalized geometric models incorporating multiple specialized models were recently introduced in [13]–[15]. However, due to the lack of preciseness in reflection methods, and the vast variation of room geometries available, an exact estimation of the RTF based on geometrical properties remain unresolved.

Due to the inefficiency of existing RTF models, alternative equalization techniques tend to measure the RTF at a finite set of points which are later incorporated to the sound processing algorithm directly [16]–[18]. However, as explained earlier, even a small-scale variation in source/receiver positions results in a drastic variation in the RTF [19], and therefore, the above method only gives accurate results at the design points, while the performance degradation present elsewhere is too significant. Additional limitations are caused by the inaccuracies involved with the point-point RTF measurements. Recent work on improving the RTF measurement via modified source and

<sup>1</sup>At high frequencies, the computational cost is too high due to the increased number of points (small wavelengths).

receiver directivity patterns include [20]–[23].

A complete equalization solution that is robust to receiver point variations was first proposed in [24] for 2D applications, which exploits a novel RTF model based on the harmonic solution to the wave equation. This model parameterizes the RTF between a fixed source and any arbitrary point within a source-free receiver region in terms of a weighted sum of 2D basis functions, while the weights need to be separately measured. Thus, the successful extraction of a finite set of parameter weights/coefficients enables RTF characterization between a given source location and any arbitrary point within a given receiver region. However, these coefficients remain unique to the source location of interest, and therefore, the slightest variation in source positioning requires a new set of RTF parameters to be measured.

In this paper, we introduce an efficient RTF parameterization in 3D, that is robust to both receiver and source variations so that the extraction of a finite set of coefficients is sufficient to characterize an entire room enclosure of interest. In other words, we derive a 3D model, which characterizes the RTF between any two arbitrary points from a primary spatial region where the source(s) lie and a secondary spatial region where the receiver(s) lie. More importantly, we impose no restrictions on the geometrical configuration of the source and receiver regions and as a result, the proposed parameterization is valid for any two arbitrary points from the given room. Following [24], this parameterization is based on the harmonic solution to the wave equation, and therefore, is derived in terms of a weighted sum of 3D basis functions. Furthermore, it only requires a minimum of  $(N_s + 1)^2(N_r + 1)^2$  coefficients to characterize the RTF over an  $N_s^{\text{th}}$  order source region and an  $N_r^{\text{th}}$  order receiver region<sup>2</sup>. We also provide a practical method to extract the aforementioned coefficients, which only requires RTF measurements over a finite set of source-receiver combinations and associated numerical processing. Given the room characteristics remain stationary over time, these measurements can be obtained using a single microphone unit and a single loudspeaker unit.

The paper is structured as follows. In Sec. II, we first decompose the room response into direct and reverberant components where the former is known and the latter is unknown. The unknown reverberant component is then parameterized in terms of a weighted sum of 3D basis functions. In Sec. III, we describe a robust method to obtain the parameter weights, which only requires a finite set of RTF measurements. Finally, in Sec. IV, we demonstrate the accuracy of the proposed parameterization, by comparing it with a simulated room based on the image source model. This section also presents an error analysis performed over a broadband frequency range.

## II. PARAMETERIZATION OF THE ROOM TRANSFER FUNCTION

### A. Problem formulation

The main objective of this paper is to have an efficient parameterization for the RTF such that it is valid for variations

<sup>2</sup>Section II-B discusses how the order of a spatial soundfield is determined over a known region and a given frequency.

in the receiver position as well as in the source position. Therefore, we first define a continuous spatial region where the source(s) lie (*source region*) and a continuous spatial region where the receiver(s) lie (*receiver region*), and the new parameterization is expected to deliver the RTF between any two arbitrary points from these two regions.

For computational simplicity, we assume the receiver region named  $\eta$  to be a sphere of radius  $R_r$  centered at the origin  $\mathbf{O}$  and the source region named  $\zeta$  to be another sphere of radius  $R_s$  centered at  $\mathbf{O}_s$  (See Fig. 1). In spherical coordinates, the receiver point within  $\eta$  is denoted by  $\mathbf{x} = (x, \theta_x, \phi_x)$  and the source location within  $\zeta$  is denoted by  $\mathbf{y} = (y, \theta_y, \phi_y)$  where  $\mathbf{y} = \mathbf{y}^{(s)} + \mathbf{R}_{sr}$  with  $\mathbf{y}^{(s)} = (y^{(s)}, \theta_y^{(s)}, \phi_y^{(s)})$  representing the same source location with respect to  $\mathbf{O}_s$  and  $\mathbf{R}_{sr}$  representing the vector connecting  $\mathbf{O}$  to  $\mathbf{O}_s$ .

In a reverberant environment, the acoustic transfer function between  $\mathbf{x}$  and  $\mathbf{y}$  can be decomposed into a direct path field and a reflected field as

$$H(\mathbf{x}, \mathbf{y}, k) = H_{\text{dir}}(\mathbf{x}, \mathbf{y}, k) + H_{\text{rvb}}(\mathbf{x}, \mathbf{y}, k) \quad (1)$$

where  $k = 2\pi f/c$  is the wave number,  $f$  is the frequency and  $c$  is the speed of sound propagation. The direct field component due to a unit amplitude point source at  $\mathbf{y}$  is independent of the room characteristics and can be given in terms of [25]

$$H_{\text{dir}}(\mathbf{x}, \mathbf{y}, k) = \frac{e^{ik\|\mathbf{x}-\mathbf{y}\|}}{4\pi\|\mathbf{x}-\mathbf{y}\|}. \quad (2)$$

However,  $H_{\text{rvb}}(\mathbf{x}, \mathbf{y}, k)$ , the corresponding reflected field incident at  $\eta$  is unknown, and completely dependent on the room characteristics. Our aim is to parameterize this unknown field so that a finite set of weights/coefficients unique to the room will be capable of predicting  $H_{\text{rvb}}(\mathbf{x}, \mathbf{y}, k)$  between any two points from  $\zeta$  and  $\eta$ .

We base our parameterization approach on the fact that the unknown  $H_{\text{rvb}}(\mathbf{x}, \mathbf{y}, k)$  incident on  $\eta$  is caused by the outward propagating wavefield from  $\zeta$ . Since both these incoming and outgoing soundfields can be represented in terms of modal decompositions<sup>3</sup>,  $H_{\text{rvb}}(\mathbf{x}, \mathbf{y}, k)$  could also be represented in terms of a similar decomposition. The coefficients of such a decomposition will then enable the user to predict the RTF between two arbitrary points from  $\zeta$  and  $\eta$ . Following the above concept, we first decompose the reverberant field at  $\eta$  due to an arbitrary outgoing field from  $\zeta$  and then derive an exact decomposition for the room transfer function.

### B. Modal decomposition of an arbitrary reverberant field

Consider an arbitrary outgoing field from  $\zeta$ , which can be represented in terms of a spherical harmonic decomposition<sup>4</sup> with respect to  $\mathbf{O}_s$  as

$$S_{\text{out}}(\mathbf{z}^{(s)}, k) = \sum_{n=0}^{N_s} \sum_{m=-n}^n \beta_{nm}^{(s)}(k) h_n(kz^{(s)}) Y_{nm}(\theta_z^{(s)}, \phi_z^{(s)}) \quad (3)$$

<sup>3</sup>A decomposition using the basis functions of the solution to the wave equation.

<sup>4</sup>Other coordinate systems could be used instead of spherical coordinates, resulting in a different set of basis functions.

where  $\mathbf{z}^{(s)} = (z^{(s)}, \theta_z^{(s)}, \phi_z^{(s)})$  denotes the observation point outside of  $\zeta$ ,  $\beta_{nm}^{(s)}(k)$  denotes the coefficients of the outgoing soundfield caused by the source distribution in  $\zeta$ ,  $Y_{nm}(\theta, \phi)$  denotes the spherical harmonic of order  $n$  and degree  $m$ ,  $h_n(\cdot)$  represents the spherical Hankel function of the first kind with order  $n$  and  $N_s = \lceil keR_s/2 \rceil$  denotes the exterior field truncation limit for a source distribution with its furthest source located at a distance of  $R_s$  [26]<sup>5</sup>.

If the resulting reflected field at  $\eta$  due to each unit amplitude outgoing mode of (3) can be extracted, the total reflected field caused by an arbitrary outgoing field can be successfully predicted. To demonstrate the above statement, let's consider a unit amplitude outgoing wave of order  $n'$  and mode  $m'$

$$\beta_{nm}^{(s)}(k) = \begin{cases} 1, & n = n' \text{ and } m = m' \\ 0, & \text{otherwise,} \end{cases} \quad (4)$$

producing

$$S_{\text{out}}(\mathbf{z}^{(s)}, k) = h_{n'}(kz^{(s)})Y_{n'm'}(\theta_z^{(s)}, \phi_z^{(s)}). \quad (5)$$

For this particular outgoing source field, there will be a resulting reflected field present at the receiver region  $\eta$ . Irrespective of the geometrical configuration of  $\zeta$  and  $\eta$ , the mirror images of the sources within  $\zeta$  are always outside of  $\eta$  and therefore, the aforementioned reflected field will be a source free incoming field. Such a soundfield can be given in terms of a harmonic decomposition of the form

$$R_{n'm'}(\mathbf{x}, k) = \sum_{v=0}^{N_r} \sum_{\mu=-v}^v \alpha_{v\mu}^{n'm'}(k) j_v(kx) Y_{v\mu}(\theta_x, \phi_x) \quad (6)$$

where  $\alpha_{v\mu}^{n'm'}(k)$  denotes the soundfield coefficients of the reverberant field incident at  $\eta$  caused by an unit amplitude  $n^{\text{th}}$  order and  $m^{\text{th}}$  mode outgoing soundfield at  $\zeta$ ,  $j_n(\cdot)$  represents the spherical Bessel function of order  $n$  and  $N_r = \lceil keR_r/2 \rceil$  denotes the interior field truncation limit [26]<sup>6</sup>. If  $\alpha_{v\mu}^{n'm'}(k)$  of (6) can be recorded up to order  $N_r$  for each unit amplitude outgoing mode from  $\zeta$ , the reverberant field at  $\eta$  due to an arbitrary outgoing field at  $\zeta$  can be derived using (3), (5) and (6) as

$$P_{\text{rb}}(\mathbf{x}, k) = \sum_{n=0}^{N_s} \sum_{m=-n}^n \sum_{v=0}^{N_r} \sum_{\mu=-v}^v \beta_{nm}^{(s)}(k) \alpha_{v\mu}^{nm}(k) j_v(kx) Y_{v\mu}(\theta_x, \phi_x). \quad (7)$$

<sup>5</sup>The truncation of a spherical harmonic based soundfield decomposition was originally derived based on the high pass behavior of Bessel functions. More precisely, Bessel functions of the form  $j_n(x)$  at  $x \leq kr$  tend to be close to zero for orders above  $N = ker/2$ , and play an insignificant role in the infinite summation. In case the reader is confused by the absence of Bessel functions in (3), please note that the modal coefficients  $\beta_{nm}^{(s)}(k)$  of any arbitrary outgoing soundfield can be represented in terms of Bessel functions [26].

<sup>6</sup>Truncation is derived following the same principle discussed earlier.

### C. Modal decomposition of the room transfer function

Now consider a unit amplitude point source at  $\mathbf{y}^{(s)} \in \zeta$ , producing outgoing soundfield coefficients  $\beta_{nm}^{(s)}(k)$  of the form [25]

$$\beta_{nm}^{(s)}(k) = ikj_n(ky^{(s)})Y_{nm}^*(\theta_y^{(s)}, \phi_y^{(s)}). \quad (8)$$

The corresponding reflected field at  $\eta$  describes the unknown reverberant component  $H_{\text{rb}}(\mathbf{x}, \mathbf{y}, k)$  of (1). This can be derived using (7) and (8) as

$$H_{\text{rb}}(\mathbf{x}, \mathbf{y}, k) = ik \sum_{n=0}^{N_s} \sum_{m=-n}^n \sum_{v=0}^{N_r} \sum_{\mu=-v}^v \alpha_{v\mu}^{nm}(k) j_n(ky^{(s)}) j_v(kx) Y_{nm}^*(\theta_y^{(s)}, \phi_y^{(s)}) Y_{v\mu}(\theta_x, \phi_x). \quad (9)$$

Therefore, the total acoustic transfer function between any two arbitrary points from the source region  $\zeta$  and the receiver region  $\eta$  can be given in terms of the direct field (2) and reflected field (9) components as

$$H(\mathbf{x}, \mathbf{y}, k) = \frac{e^{ik\|\mathbf{x}-\mathbf{y}\|}}{4\pi\|\mathbf{x}-\mathbf{y}\|} + ik \sum_{n=0}^{N_s} \sum_{m=-n}^n \sum_{v=0}^{N_r} \sum_{\mu=-v}^v \alpha_{v\mu}^{nm}(k) j_n(ky^{(s)}) j_v(kx) Y_{nm}^*(\theta_y^{(s)}, \phi_y^{(s)}) Y_{v\mu}(\theta_x, \phi_x). \quad (10)$$

*Comments:*

- Based on the above result (10), the RTF can be parameterized in terms of a spherical harmonic decomposition. If  $\alpha_{v\mu}^{nm}(k)$ , the weights/coefficients of this parameterization can be accurately captured, they can be used to derive the RTF between any two arbitrary points from a continuous spatial region where the source(s) lie and a continuous spatial region where the receiver(s) lie.
- To generalize the RTF over an  $N_s^{\text{th}}$  order source region  $\zeta$ , where  $N_s = k_{\text{max}}eR_s/2$  and an  $N_r^{\text{th}}$  order receiver region  $\eta$ , where  $N_r = k_{\text{max}}eR_r/2$ , the above parameterization requires a minimum of  $(N_r + 1)^2(N_s + 1)^2$  unique coefficients of the form  $\alpha_{v\mu}^{nm}(k)$ . For example, when the maximum frequency of interest is  $f_{\text{max}} = 1$  kHz and the source and receiver regions of interest are both spheres of radius 0.2 m with  $N_s = N_r = 5$ , a fixed number of 1296 unique coefficients are required to calculate the RTF between any two arbitrary points  $\mathbf{x}$  and  $\mathbf{y}$  from  $\eta$  and  $\zeta$  respectively. In broadband applications, the total coefficient count will increase with each frequency sample  $f_o$  requiring an additional set of  $(k_o e R_r / 2 + 1)^2 (k_o e R_s / 2 + 1)^2$  coefficients.
- Due to the decomposition of direct and reverberant components, this parameterization supports any configuration of  $\eta$  and  $\zeta$ . As shown in Fig. 1 they can be either completely separated from each other

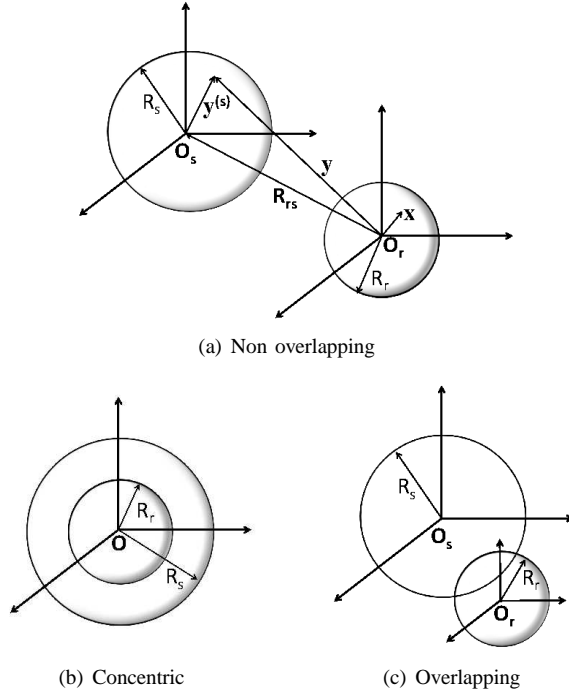


Fig. 1. Different configurations of the source region  $\zeta$  and the receiver region  $\eta$

with  $\|\mathbf{R}_{sr}\| > (R_r + R_s)$  (Fig. 1(a)), concentric with  $\|\mathbf{R}_{sr}\| = 0$  (Fig. 1(b)) or overlapping on each other with  $\|\mathbf{R}_{sr}\| < (R_r + R_s)$  (Fig. 1(c)). Therefore, (10) can be used to either partially or fully characterize the room according to user requirements.

- Taking all of the above properties into consideration, the proposed parameterization can be interpreted as a modal based solution to the wave equation in arbitrary room environments. Compared to the classical mode solution to the RTF defined for rectangular rooms [1], this parameterization has three main advantages. First, this method is applicable to any arbitrary room geometry. Second, practical room environments (having furniture etc.) with irregular boundary conditions are extremely difficult to be characterized by the classical solution, whereas the new method is valid for any arbitrary acoustic environment. Third, unlike the total mode count ( $\frac{4\pi}{3}V(\frac{f}{c})$  where  $V$  denotes room volume) of the classical model, that of the new model ( $(N_r + 1)^2(N_s + 1)^2$ ) can be reduced by defining smaller  $R_s$  and  $R_r$  values to improve the computational efficiency. (This property is specially advantageous at the Schroeder frequency [27] where the classical model will require a very large mode count resulting in a Gaussian distribution.

### III. ESTIMATION OF ROOM TRANSFER FUNCTION COEFFICIENTS

In this section, we present the procedure of estimating the RTF coefficients  $\alpha_{v\mu}^{nm}(k)$  of (10) for a pre-defined source region and a pre-defined receiver region. As explained earlier,  $\alpha_{v\mu}^{nm}(k)$  represents the  $v^{\text{th}}$  order and  $\mu^{\text{th}}$  mode reverberant field

coefficient within  $\eta$  caused by a unit amplitude  $n^{\text{th}}$  order and  $m^{\text{th}}$  mode outgoing wavefront originated at  $\zeta$  (5). For each outgoing mode from  $\zeta$ , there will be  $(N_r + 1)^2$  number of unique coefficients describing the reverberant field incident at  $\eta$ , and to generalize an  $N_s^{\text{th}}$  order source region, a total number of at least  $(N_s + 1)^2(N_r + 1)^2$  coefficients needs to be extracted.

It is important to note that, in practice, the following method does not require the physical production of unit amplitude outgoing modes from  $\zeta$  and associated room response recordings, but only requires the acquisition of room response between a set of receivers distributed within  $\eta$  and a set of loudspeakers distributed within  $\zeta$ , each transmitting a unit amplitude signal. Furthermore, given the room characteristics remain stationary over time, these measurements can be obtained using a single microphone unit and a single loudspeaker unit. However, for the purpose of deriving this result, we will discuss a method to generate unit amplitude modal wavefronts propagating outward from the source region, and a soundfield recording technique to extract the corresponding room responses. Theoretically, these processes are required to be repeated for a minimum of  $(N_s + 1)^2$  number of different cases, but their physical implementation will be proven to be needless in sec. III-B1.

#### A. Synthesis of a unit amplitude outgoing mode originated from the source region

Let us first consider the problem of producing a unit amplitude outgoing mode from  $\zeta$  with respect to  $\mathbf{O}_s$  (5). In order to account for all the significant outgoing modes from  $\zeta$ ,  $n'$  and  $m'$  from (5) has to be varied from 0 to  $N_s$  and from  $-n$  to  $n$  respectively. This results in a total number of  $(N_s + 1)^2$  distinct soundfield production cases and corresponding weight vectors.

For each case, we propose a mode matching approach where the modal coefficients of the desired outgoing field (4) are matched with those of the outgoing wavefield produced by an array of loudspeakers distributed within  $\zeta$ . Consider  $L$  number of point sources arbitrarily distributed within  $\zeta$ , where the  $\ell^{\text{th}}$  source ( $\ell = 1 \dots L$ ) is located at  $\mathbf{y}_{\ell}^{(s)} = (y_{\ell}^{(s)}, \theta_{y_{\ell}}^{(s)}, \phi_{y_{\ell}}^{(s)})$  with respect to  $\mathbf{O}_s$ . The weighted sum of loudspeaker outputs will produce an outgoing soundfield of the form (3) where  $\beta_{nm}^{(s)}(k)$  is [26]

$$\beta_{nm}^{(s)}(k) = \sum_{\ell=1}^L w_{\ell}(k) i k j_n(k y_{\ell}^{(s)}) Y_{nm}^*(\theta_{y_{\ell}}^{(s)}, \phi_{y_{\ell}}^{(s)}) \quad (11)$$

with  $w_{\ell}(k)$  representing the weights at each point source. Our objective is to derive loudspeaker weights that will produce (4), a unit amplitude outgoing wave of order  $n'$  and mode  $m'$ .

This can be achieved by equating (4) to (11), which forms a set of linear equations of the form

$$\mathbf{T} \mathbf{w}^{n'm'} = \beta^{n'm'} \quad (12)$$

where

$$\mathbf{T} = i k \begin{bmatrix} t_{00}(k, \mathbf{y}_1^{(s)}) & \dots & \dots & \dots & t_{00}(k, \mathbf{y}_L^{(s)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{N_s N_s}(k, \mathbf{y}_1^{(s)}) & \dots & \dots & \dots & t_{N_s N_s}(k, \mathbf{y}_L^{(s)}) \end{bmatrix} \quad (13)$$

is an  $(N_s + 1)^2 \times L$  translation matrix with  $t_{nm}(k, \mathbf{y}^{(s)}_\ell) = j_n(ky_\ell^{(s)})Y_{nm}^*(\theta_{y_\ell}^{(s)}, \phi_{y_\ell}^{(s)})$ ,  $\mathbf{w}^{n'm'} = [w_1^{n'm'}(k) \dots w_L^{n'm'}(k)]^T$  is an  $L$  long vector of loudspeaker weights and  $\boldsymbol{\beta}^{n'm'} = [0 \dots 0 \ 1 \ 0 \dots 0]^T$  is an  $(N_s + 1)^2$  long vector of desired field coefficients where the  $(n'^2 + n' + m' + 1)^{\text{th}}$  element is 1 while all others are zero. Since  $\mathbf{T}$  and  $\boldsymbol{\beta}^{n'm'}$  are both known, the required weights at each loudspeaker can be solved using

$$\mathbf{w}_{n'm'} = \mathbf{T}^\dagger \boldsymbol{\beta}^{n'm'} \quad (14)$$

where  $\mathbf{T}^\dagger$  denotes the pseudoinverse. To avoid spatial aliasing,

$$L \geq (N_s + 1)^2 \quad (15)$$

has to be satisfied with (14) yielding the minimum energy weight solution.

While (14) provides a numerical solution to the required loudspeaker weights, it is also important to decide upon a robust array geometry. The spherical geometry has been widely used for rendering and acquisition of spatial soundfields [28]–[30] however, its performance in the above task can be predicted to be less robust due to the Bessel functions present in  $\mathbf{T}$ . When  $j_n(ky_\ell^{(s)})$  of (13) approaches zero crossings, the condition number of  $\mathbf{T}$  increases<sup>7</sup>, and since the pseudoinverse of an ill-conditioned matrix is often erroneous, those errors will be propagated to the weight solution in (14). Similar issues were experienced in spherical microphone arrays used for interior field recording [28], which were later overcome by using rigid arrays [30] or variable radii arrays like the spherical shell array given in [31] and the dual spherical array given in [32]. For the loudspeaker array of interest, a rigid geometry requires the incorporation of scattering effects and a dual spherical array requires twice the number of loudspeakers. Therefore, in this paper, we opt for the simplest geometry of choice, an open spherical shell. A spherical shell array is equally distributed in the angular space while the distance to each loudspeaker  $y_\ell^{(s)}$  randomly varies (with a uniform distribution) between a virtual spherical shell of outer radius  $R_s$  and an inner radius  $R'_s$ . For in depth reasoning and additional solution to the open sphere inverse problem, the reader may refer to [33].

An alternate approach for achieving robust loudspeaker arrays was recently introduced in [34] for 2D soundfields, and in [35] for 3D soundfields where the conventional circular/spherical arrays of monopole loudspeakers were replaced by those of higher order loudspeakers. A 3D higher order loudspeaker of order  $D$  is capable of producing polar responses up to the  $D^{\text{th}}$  order. This solution significantly reduces the minimum requirement of loudspeaker units by a factor of  $1/(D + 1)^2$  at the expense of increased complexity at each loudspeaker unit and therefore, it is more suitable for sound reproduction in large spatial areas. Since the practical implementation of higher order loudspeakers are still in the design stage, the above approach is not used in this paper. However, the reader is encouraged to refer to [34], [35] for a detailed description of the array processing involved with sound rendering using higher order loudspeakers.

<sup>7</sup>The 2-norm condition number of a matrix  $\mathbf{T}$  is defined by  $\kappa_2(\mathbf{T}) = \|\mathbf{T}\|_2 \cdot \|\mathbf{T}^\dagger\|_2$  and for a well conditioned matrix, it will be close to 1.

## B. Extracting the room response at the receiver region

Once the desired outgoing waves are synthesized at  $\zeta$ , the next step involves the extraction of resulting room reflections incident at  $\eta$ . It is important to note that all recordings obtained at  $\eta$  carry both direct and reflected wavefronts originated at  $\zeta$ , and since we are only parameterizing the reverberant field, the direct field component at each sensor output must be removed prior to further processing.

Furthermore, as the reverberant field of interest (6) is a source free incoming field, its extraction can be treated as an *interior field recording problem*. The conventional approach to record an  $N_r^{\text{th}}$  order incoming spatial soundfield requires a minimum of  $(N_r + 1)^2$  omnidirectional microphones equally distributed on a spherical surface enclosing the region of interest [29]. However, this approach has been proven to be less robust due to the above mentioned "Bessel zero problem" and as explained earlier, alternate geometries were later proposed in [30]–[32] to overcome this issue.

A further improved solution to the interior field recording problem was recently introduced in [26] where the omnidirectional microphones were replaced by higher order (HO) microphones. A higher order microphone of order  $A$  is capable of recording an  $A^{\text{th}}$  order spatial soundfield with respect to the microphone's local origin. Thus, the use of  $A^{\text{th}}$  order microphones in recording an  $N_r^{\text{th}}$  order soundfield substantially reduces the minimum requirement of measurements by a factor of  $1/(A + 1)^2$  at the expense of added complexity at each microphone unit. Compared to the conventional omnidirectional microphone array, this approach also showed a significant improvement in the condition number of the translation matrix, which in turn increased the array's robustness. Furthermore, unlike the higher order loudspeakers, the practical implementation of HO microphones are relatively simple and there exist a number of different designs that were successfully implemented in practice [36]. The "Eigenmike" is one such commercially available fourth order microphone with an active frequency range of 0 – 6.5 kHz.

Due to the aforementioned efficiency and availability of HO microphones, we propose an array of  $Q$  identical  $A^{\text{th}}$  order microphones to be employed in the coefficient extraction process. For each unit amplitude outgoing wavefield produced at  $\zeta$ , there will be an  $N_r^{\text{th}}$  order soundfield incident at  $\eta$ , and according to [26], the extraction of such a soundfield requires a minimum of  $Q \geq (N_r + 1)^2 / (A + 1)^2$  HO microphone units distributed in any arbitrary geometry enclosing the region of interest. The translation between the HO microphone outputs and the desired reverberant soundfield is based on a coefficient translation theorem developed in [26], which will be discussed in detail in sec. III-B3.

*1) Higher order microphone:* Let us now briefly discuss the functionality of a 3D HO microphone. Consider the  $q^{\text{th}}$  ( $q = 1 \dots Q$ ) HO microphone located at  $\mathbf{O}_q$  with  $\mathbf{R}_q = (R_q, \theta_q, \phi_q)$  representing the vector connecting  $\mathbf{O}$  to  $\mathbf{O}_q$ . For numerical simplicity, we assume it is designed following the open array geometry given in [29], where an  $A^{\text{th}}$  order microphone is composed of an array of  $Q' \geq (A + 1)^2$  number

of omnidirectional microphones equally distributed along a virtual spherical surface of radius  $r = Ac/\pi e f_{\max}$  where  $f_{\max}$  denotes the maximum frequency of interest in broadband operation. The  $q^{\text{th}}$  ( $q' = 1 \dots Q'$ ) microphone of this array will be located at  $\mathbf{r}_{qq'} = (r_{qq'}, \theta_{qq'}, \phi_{qq'})$  with respect to  $\mathbf{O}_q$  recording

$$P_{q'}^{(q)}(\mathbf{r}_{qq'}, k) = \sum_{a=0}^A \sum_{b=-a}^a \gamma_{ab}^{(q)}(k) j_a(kr_{qq'}) Y_{ab}(\theta_{qq'}, \phi_{qq'}) \quad (16)$$

where  $r_{qq'} = r$  for all  $q'$  and  $\gamma_{ab}^{(q)}(k)$  represents the soundfield coefficients with respect to  $\mathbf{O}_q$ . Over the entire array, there will be a total of  $Q'$  recordings of the above form, and based on the orthogonal property of spherical harmonics [29], they can be collectively combined to extract  $\gamma_{ab}^{(q)}(k)$  using

$$\gamma_{ab}^{(q)}(k) = \frac{1}{j_a(kr_{qq'})} \sum_{q'=1}^{Q'} P_{q'}^{(q)}(\mathbf{r}_{qq'}, k) Y_{ab}^*(\theta_{qq'}, \phi_{qq'}). \quad (17)$$

When the above microphone is used to record the room response caused by the  $n^{\text{th}}$  order and  $m^{\text{th}}$  mode unit outgoing wavefield originated from  $\zeta$ , the incident pressure at the  $q^{\text{th}}$  omnidirectional microphone will be

$$P_{q'}^{(q)}(n, m, \mathbf{r}_{qq'}, k) = \sum_{\ell=1}^L w_{\ell}^{nm}(k) H(k, \mathbf{x}_{q'}^{(q)}, \mathbf{y}_{\ell}) \quad (18)$$

where  $H(k, \mathbf{x}_{q'}^{(q)}, \mathbf{y}_{\ell})$  denotes the RTF between the omnidirectional receiver at  $\mathbf{x}_{q'}^{(q)} = \mathbf{R}_q + \mathbf{r}_{qq'}$  and the weighted point source at  $\mathbf{y}_{\ell} = \mathbf{y}_{\ell}^{(s)} + \mathbf{R}_{sr}$  with respect to  $\mathbf{O}$ . Substituting for (17) from (18) we derive the corresponding outputs at the  $q^{\text{th}}$  HO microphone as

$$\gamma_{ab}^{(q,n,m)}(k) = \sum_{\ell=1}^L w_{\ell}^{nm}(k) \underbrace{\frac{1}{j_a(kr_{qq'})} \sum_{q'=1}^{Q'} H(k, \mathbf{x}_{q'}^{(q)}, \mathbf{y}_{\ell}) Y_{ab}^*(\theta_{qq'}, \phi_{qq'})}_{\tilde{\gamma}_{ab}^{(q,\ell)}(k)} \quad (19)$$

where  $\tilde{\gamma}_{ab}^{(q,\ell)}(k)$  denotes the incident soundfield coefficients at  $\mathbf{O}_q$  caused by a unit amplitude loudspeaker located at  $\mathbf{y}_{\ell}$ .

Consequently, if  $\tilde{\gamma}_{ab}^{(q,\ell)}(k)$ , the room response between the  $\ell^{\text{th}}$  loudspeaker and the  $q^{\text{th}}$  HO microphone can be recorded for all  $L$  loudspeakers and all  $Q$  HO microphones, then,  $\gamma_{ab}^{(q,n,m)}(k)$  can be easily derived using the linearity property given in (19). This profound result significantly simplifies the coefficient extraction process by completely eliminating the requirement for (12)'s physical implementation. Furthermore, all  $(N_s + 1)^2$  distinct cases of (12) can now be synthesized using the same set of  $\tilde{\gamma}_{ab}^{(q,\ell)}(k)$  measurements and appropriate numerical processing.

2) *Removal of the direct soundfield:* As  $\gamma_{ab}^{(q,n,m)}(k)$  are considered as the HO microphone outputs, it is essential to remove their direct path components prior to further array processing. The coefficients  $\gamma_{ab}^{(q,n,m)}(k)$  can be decomposed into direct and reverberant field components as

$$\gamma_{ab}^{(q,n,m)}(k) = \gamma_{ab(\text{dir})}^{(q,n,m)}(k) + \gamma_{ab(\text{rvb})}^{(q,n,m)}(k) \quad (20)$$

where the direct field component is known to be [25]

$$\gamma_{ab(\text{dir})}^{(q,n,m)}(k) = \sum_{\ell=1}^L w_{\ell}^{(nm)}(k) ikh_a(kR_{q\ell}) Y_{ab}^*(\theta_{q\ell}, \phi_{q\ell}) \quad (21)$$

with  $(R_{q\ell}, \theta_{q\ell}, \phi_{q\ell})$  denoting the spherical coordinates of  $\mathbf{R}_{q\ell} = \mathbf{y}_{\ell} - \mathbf{R}_q$ . Therefore, once  $\gamma_{ab}^{(q,n,m)}(k)$  are obtained, (20) and (21) can be used to eliminate their direct field components.

3) *Array of higher order microphones:* The final step in array processing involves the translation of  $\gamma_{ab(\text{rvb})}^{(q,n,m)}(k)$  to the desired RTF coefficients  $\alpha_{v\mu}^{nm}(k)$ . As mentioned earlier, this can be done following the coefficient translation theorem introduced in [35] as

$$\gamma_{ab(\text{rvb})}^{(q,n,m)}(k) = \sum_{v=0}^{N_r} \sum_{\mu=-v}^v \alpha_{v\mu}^{nm}(k) \hat{S}_{va}^{\mu b}(\mathbf{R}_q) \quad (22)$$

where

$$\hat{S}_{va}^{\mu b} = 4\pi i^{a-v} \sum_{l=0}^{\infty} i^l (-1)^{2\mu-b} j_l(kR_q) Y_{l(b-\mu)}^*(\theta_q, \phi_q) \times \sqrt{(2v+1)(2a+1)(2l+1)/4\pi W_1 W_2}, \text{ with}$$

$$W_1 = \begin{pmatrix} v & a & l \\ 0 & 0 & 0 \end{pmatrix} \text{ and } W_2 = \begin{pmatrix} v & a & l \\ \mu & -b & (b-\mu) \end{pmatrix}$$

representing Wigner 3-j symbols. For all  $Q$  number of microphones, (22) can be interpreted in matrix form as,

$$\boldsymbol{\gamma} = \mathbf{T}' \boldsymbol{\alpha} \quad (23)$$

where  $\boldsymbol{\gamma} = [\gamma_{00}^{(1,n,m)}, \dots, \gamma_{AA}^{(1,n,m)}, \dots, \gamma_{00}^{(Q,n,m)}, \dots, \gamma_{AA}^{(Q,n,m)}]^T$  is a  $Q(A+1)^2$  long vector,  $\boldsymbol{\alpha} = [\alpha_{00}, \dots, \alpha_{N_r N_r}]^T$  is a  $(N_r+1)^2$  long vector, and

$$\mathbf{T}' = \begin{bmatrix} \hat{S}_{00}^{00}(\mathbf{R}_1) & \dots & \dots & \dots & \hat{S}_{N_r 0}^{N_r 0}(\mathbf{R}_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{S}_{0A}^{0A}(\mathbf{R}_Q) & \dots & \dots & \dots & \hat{S}_{N_r A}^{N_r A}(\mathbf{R}_Q) \end{bmatrix} \quad (24)$$

is a  $Q(A+1)^2 \times (N_r+1)^2$  matrix. As  $\mathbf{T}'$  is known, and the local recordings in  $\boldsymbol{\gamma}$  can be derived from (19), (23) can be solved to find the desired coefficients  $\boldsymbol{\alpha}$ , using

$$\boldsymbol{\alpha} = \mathbf{T}'^{\dagger} \boldsymbol{\gamma}. \quad (25)$$

To avoid spatial aliasing,

$$Q \geq (N_r + 1)^2 / (A + 1)^2 \quad (26)$$

has to be satisfied [26] with (25) yielding a least squares solution.

### C. Summary of the coefficient extraction process

The coefficient extraction process involved with the RTF parameterization proposed over an  $N_s^{\text{th}}$  order source region and an  $N_r^{\text{th}}$  order receiver region is summarized as follows. Theoretically, this process requires  $(N_s + 1)^2$  number of distinct outgoing waves created at  $\zeta$  and the same number of reverberant field extractions simultaneously carried out at  $\eta$ . Each case requires a minimum of  $(N_s + 1)^2$  point sources or  $(N_s + 1)^2 / (D + 1)^2$  number of  $D^{\text{th}}$  order loudspeakers at  $\zeta$  to synthesize the outgoing field, and a minimum of  $(N_r + 1)^2$  omnidirectional microphones or  $(N_r + 1)^2 / (A + 1)^2$  number of  $A^{\text{th}}$  order microphones to extract the reverberant field at  $\eta$ . Depending on the size and frequency content of the source and receiver regions, the user can employ any combination of point sources, higher order sources, omnidirectional microphones and higher order microphones. As explained in sections III-A and III-B, this work illustrates one of the above combinations, an open spherical shell array of point sources and an open spherical array of HO microphones.

However, in practice, it is only required to extract the room response between each loudspeaker and each microphone from the above arrays, and by incorporating these measurements with the numerical computations given in (14), (19), (20) and (25), the desired RTF coefficients can be successfully derived. Moreover, given the room characteristics remain stationary over time, the above measurements can be obtained using a single microphone unit and a single loudspeaker unit moved along the respective arrays.

Practical limitations of the proposed measurement method arise with large  $R_s$  and  $R_r$  values at high frequencies due to the increased number of modal components required to describe the spatial soundfield of interest. Common forms of these limitations include the large requirement of microphone and loudspeaker numbers in non-stationary conditions, increased demand for high computational power, and the design and implementation constraints involved with large spherical/shell geometries. Furthermore, the proposed use of HO microphones to record  $\eta$  may not be practically feasible at present due to the lack of affordable HO microphones that are commercially available. The above constraints may however be overcome by defining smaller source and receiver regions to suit the application of interest, and the HO microphone array can be easily replaced with omnidirectional ones to reduce costs. In addition to the above constraints, in a real room, temperature (and to a lesser extent humidity) fluctuations can cause the room impulse responses to change, especially in the late reverberant tails and the higher frequency components. However, the proposed approach is likely to be accurate at low frequencies and to the modeling the RTF components of low order reflections.

### D. Approximate parameterization error

The total error involved with the proposed RTF parameterization has several components that will be encountered at different stages of the parameterization process. The first component will appear at the loudspeaker array processing phase (12), in the forms of truncation error ( $N_s = \lceil keR_s/2 \rceil$ )

in (3), and least squares error in (14) related to the geometry and numbers of loudspeakers. The next component will occur at each HO microphone, again in the forms of truncation error ( $A = \lceil keR_r/2 \rceil$ ) in (16), and Bessel-zero error in (17). The final component will develop in the coefficient translation phase, once more in the forms of truncation error ( $N_r = \lceil keR_r/2 \rceil$ ) in (6), and least squares error in (25). A detailed decomposition of each of the above components is of least interest in the current context, thus, we only study the total error. For computational simplicity, we define an approximate error averaged over a finite number of design points from  $\zeta$  and  $\eta$  as

$$E = \frac{\sum_{g=1}^G \|\tilde{H}(\mathbf{x}_g, \mathbf{y}_g, k) - H(\mathbf{x}_g, \mathbf{y}_g, k)\|}{\sum_{g=1}^G \|\tilde{H}(\mathbf{x}_g, \mathbf{y}_g, k)\|} \quad (27)$$

where  $G$  denotes the number of source and receiver point combinations being considered and  $\tilde{H}$  denotes the existing RTF.

## IV. SIMULATIONS

In the following simulation examples, we illustrate the accuracy of the proposed RTF parameterization in broadband applications. A  $6 \times 5 \times 2.5$  m rectangular room was considered as the reverberant environment with its center defined as the origin  $\mathbf{O}$ . The RTF was parameterized over a spherical receiver region  $\eta$  of radius  $R_r = 0.4$  m centered about  $\mathbf{O}$  and a spherical source region  $\zeta$  of radius  $R_s = 0.4$  m centered about  $\mathbf{O}_s$ . The location of  $\mathbf{O}_s$  was varied accordingly to simulate a non-overlapping and an overlapping configuration of  $\zeta$  and  $\eta$ . The design frequency range was assumed up to  $f_{\text{max}} = 1$  kHz producing a tenth order receiver region ( $N_{r(\text{max})} = 10$ ) and a tenth order source region ( $N_{s(\text{max})} = 10$ ). For frequencies below  $f_{\text{max}}$ , the truncations limits would drop, and therefore, when operating with varying frequencies,  $N_s$  and  $N_r$  were varied accordingly.

From (15), the synthesis of a unit amplitude outgoing wave from  $\zeta$  required a minimum of  $L = 121$  point sources distributed in a preferred geometry. As explained in Sec. (III-A), we opted for a spherical shell geometry, which required the sources to be equally distributed in the angular space and randomly varied in the magnitude space. While [37] provided an approximate solution for the desired angular distribution, the distance to each source  $\|\mathbf{y}^{(s)}\|$  was randomly varied (with uniform distribution) between a spherical shell of outer radius of  $R_s = 0.4$  m and an inner radius of  $R'_s = 0.3$  m.

The reason behind selecting a spherical shell geometry over the conventional single sphere geometry was to improve the array robustness, and we validated this decision by comparing the condition number of the translation matrix  $\kappa_2(\mathbf{T})$  related to both geometries. As shown in Fig. 2, the condition number  $\kappa_2(\mathbf{T})$  was plotted against frequency for a spherical shell geometry with the above parameters and a single sphere geometry of radius  $R_s = 0.4$  m. The expected ill-conditioning of the single sphere geometry is very much evident with  $\kappa_2(\mathbf{T})$  reaching a couple of large peaks at  $f = 420$  Hz and  $f = 850$  Hz. In contrast,  $\kappa_2(\mathbf{T})$  of the spherical shell geometry gives

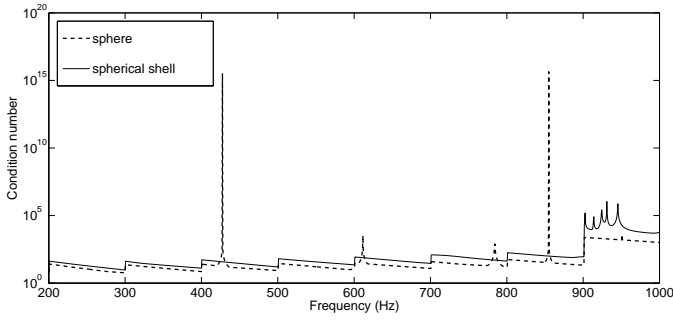


Fig. 2. Condition number variation of  $T$  in (12).

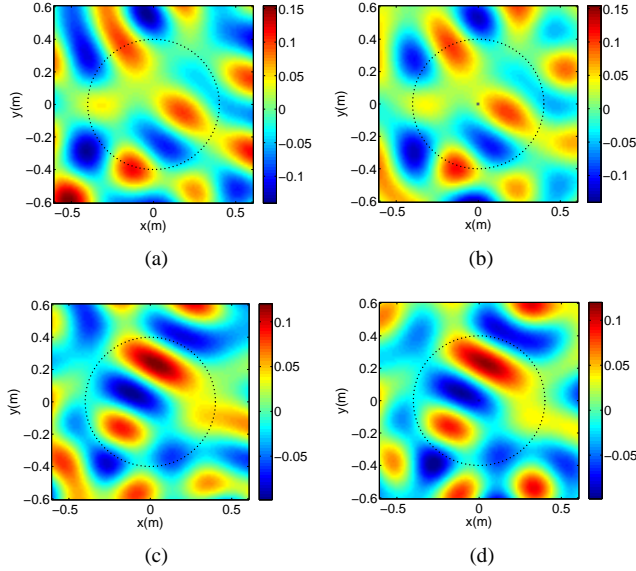


Fig. 3. Actual and reconstructed RTF between a fixed source location  $\mathbf{y}$  and all points in the receiver region for a non overlapping distribution of  $\eta$  and  $\zeta$  with  $\mathbf{R}_{sr} = (1, 1, 0.5)$  m. (a) Actual and (b) reconstructed RTF for  $\mathbf{y} = (1.05, 1.05, 0.5707)$  m. (c) Actual and (d) reconstructed RTF for  $\mathbf{y} = (1.15, 1.15, 0.6207)$  m.

much improved results by avoiding all of the above peaks. Therefore, we can conclude that a variation of  $\|\mathbf{y}^{(s)}\|$  in  $T$  successfully overcomes the Bessel zero problem in solving (14). The sawtooth characteristic of the condition number variation may have caused by the frequency-dependent mode order ( $N_s$ ).

Once the unit amplitude outgoing wavefields were synthesized at  $\zeta$ , it was required to extract the resulting reverberant fields incident at  $\eta$ . For this purpose, we opted for a spherical array geometry of radius  $R_r = 0.4$  m enclosing  $\eta$ , and according to (26), a minimum of 121 omnidirectional microphones were required to avoid spatial aliasing. To improve the array robustness and to reduce the number of measurements, we replaced the omnidirectional microphones with third order ones ( $A = 3$ ), which substantially reduced the minimum requirement of  $Q$  to  $(N_r + 1)^2 / (A + 1)^2 \approx 9$ . It is important to note that  $Q$  was restricted to square numbers in order to facilitate the spatial distribution given in [37], which provides an approximate solution to the equal division of a spherical surface.

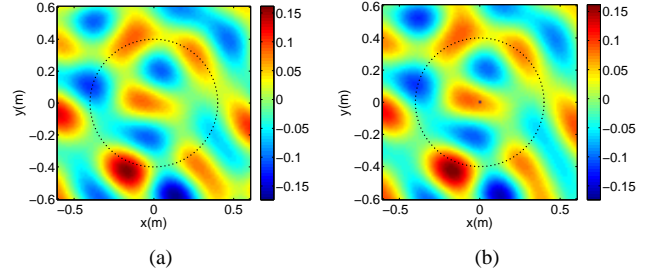


Fig. 4. (a) Actual and (b) reconstructed RTF between a fixed receiver location at  $\mathbf{y} = (0, 0, 0)$  m and all points in the source region for a non overlapping distribution of  $\eta$  and  $\zeta$  with  $\mathbf{R}_{sr} = (1, 1, 0.5)$  m.

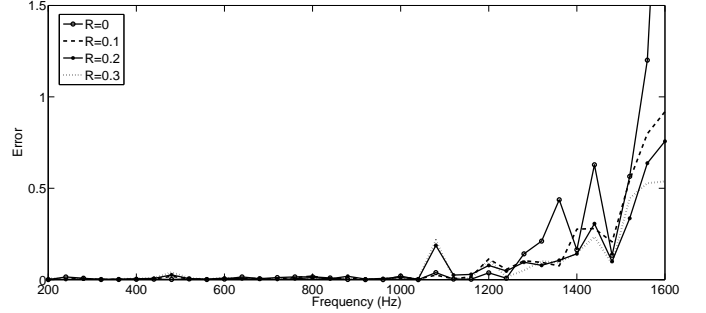


Fig. 5. Approximate parameterization error (27) for 4 different cases. In each case, the error was averaged over  $G = 7$  source and receiver point combinations, where each point was at  $[(0, 0, 0), (-R, 0, 0), (R, 0, 0), (0, -R, 0), (0, R, 0), (0, 0, -R), (0, 0, R),]$  with respect to  $\mathbf{O}_s$  and  $\mathbf{O}$  respectively.

However, in the simulations given below, we assumed the room acoustics to be stationary which in turn required only one point source and one third order microphone to measure the room response between  $121 \times 9$  combinations of  $\mathbf{y}_\ell$  and  $\mathbf{R}_q$ . The number of measurements can be further reduced by an approximate factor of  $1/(D + 1)^2$  if the point source was replaced by a higher order loudspeaker of order  $D$ .

The actual room response measurements were simulated using the image-source method [6] which defines the RTF between  $\mathbf{x}$  and  $\mathbf{y}$  in terms of

$$H(\mathbf{x}, \mathbf{y}, k) = h_0(k \|\mathbf{x} - \mathbf{y}\|) + \sum_{i=1}^I \zeta_i h_0(k \|\mathbf{x} - \mathbf{y}_i\|) \quad (28)$$

where  $\mathbf{y}_i$  and  $\zeta_i$  are the position and accumulated wall reflection coefficient of the  $i^{\text{th}}$  image source. In this paper, we considered image sources up to the second order with wall reflection coefficients  $[0.9 \ 0.9 \ 0.9 \ 0.9 \ 0.7 \ 0.7]$ .

We first looked at a non-overlapping distribution of  $\eta$  and  $\zeta$  by defining the vector from  $\mathbf{O}$  to  $\mathbf{O}_s$  as  $\mathbf{R}_{sr} = (1 \ 1 \ 0.5)$  m. Once  $\gamma_{ab}^{(q,i)}(k)$  of (19) were measured using the simulated environment given in (28), we calculated the loudspeaker weight vector  $\mathbf{W}_{nm}$  for  $(N_s + 1)^2$  distinct cases accounting for all combinations of  $n$  and  $m$ . Afterward, they were incorporated in (19) to calculate  $\gamma_{ab}^{(q,n,m)}(k)$  which was later modified using (20) and (21) to derive the HO microphone recordings of the reverberant field,  $\gamma_{ab(\text{rvb})}^{(q,n,m)}(k)$ . Finally,  $\gamma_{ab(\text{rvb})}^{(q,n,m)}(k)$  were translated to the desired RTF coefficients ( $\alpha_{v\mu}^{nm}(k)$ ), using the coefficient translation relationship given in (22).



Even though the extracted RTF coefficients are capable of mapping each point in the source region to each point in the receiver region over 0 – 1 kHz, it is not possible to plot them all at once. Therefore, we first demonstrate the array robustness to receiver variations by plotting the RTF between a particular point in the source region and all points in the receiver region for a single frequency. Next, we generated a similar plot for a secondary source location to observe the array robustness to a slight variation in source positioning. In order to further validate this property, we finally plotted the RTF between a particular point in the receiver region and all points in the source region, accounting for all possible source variations. Please note that all spatial plots were constrained to a 2D horizontal cross section with zero elevation for ease of presentation.

Figures 3(a) and 3(b) show the actual and reconstructed RTF between a source at  $\mathbf{y} = (1.05, 1.05, 0.5707)$  m and all points in the receiver region for a frequency of  $f = 900$  Hz (the circle represents the receiver region). Similarly, Figs. 3(c) and 3(d) show the corresponding results for a secondary source at  $\mathbf{y} = (1.15, 1.15, 0.6207)$  m. The reconstructed results in both cases appear almost error-less verifying the accuracy of the proposed parameterization and its robustness to receiver variations as well as a slight variation in source positioning. Demonstrating source variations in a larger scale, Figs. 4(a) and 4(b) show the actual and reconstructed RTF between a fixed receiver at  $\mathbf{y} = (0, 0, 0)$  m and all points in the source region  $\zeta$  (the circle represents the source region). As expected, the reconstructed field is almost the same as the actual one, verifying the proposed model's robustness to source variations.

Analyzing the broadband performance of the proposed parameterization, we plotted the reproduction error (27) of the recorded RTF against frequency over a range of 200 – 1700 Hz. In order to average the error, we considered a sample set of 7 points from the source region and 7 points from the receiver region resulting in  $G = 7$  one-to-one combinations<sup>8</sup>. The source and receiver points were located at  $[(0, 0, 0), (-R, 0, 0), (R, 0, 0), (0, -R, 0), (0, R, 0), (0, 0, -R), (0, 0, R),]$  with respect to  $\mathbf{O}_s$  and  $\mathbf{O}$  respectively. Figure 5 shows the results for different values of  $R$ . The error remains very low up to the maximum frequency of interest 1 kHz, beyond which it slowly builds up. The increasing error present from 1 kHz onwards is due to spatial aliasing in both reproduction and recording phases. The low amplitude errors present within the active frequency range (0.2 – 1 kHz) are possibly stemmed from the HO microphone simulations. When a fixed  $A^{\text{th}}$  order microphone with a maximum recordable frequency  $f_{\text{max}}$  is employed to record low frequencies ( $f < f_{\text{max}}$ ),  $\gamma_{ab}$  mode will be successfully recorded only if  $f < f_{\text{act}}^{(ab)}$  where  $f_{\text{act}}^{(ab)}$  denotes the activation frequency of the  $a^{\text{th}}$  order  $b^{\text{th}}$  mode component of the soundfield of interest. In other words, at low frequencies, the soundfield modes that are actually present or activated are only up to the order  $A' = \pi f_{\text{er}}/C$ , and all modes produced

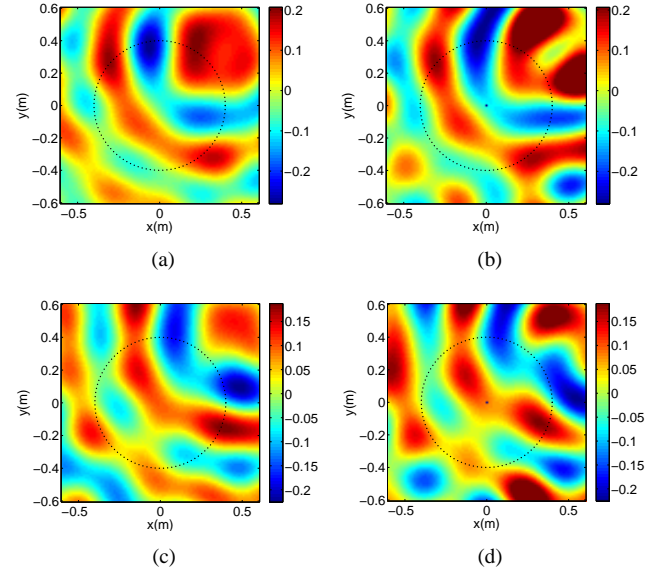


Fig. 6. Actual and reconstructed RTF between a fixed source location  $\mathbf{y}$  and all points in the receiver region for an overlapping distribution of  $\eta$  and  $\zeta$  with  $\mathbf{R}_{sr} = (0.3 \ 0.3 \ 0.3)$  m. (a) Actual and (b) reconstructed RTF for  $\mathbf{y} = (0.35, 0.35, 0.3707)$  m. (c) Actual and (d) reconstructed RTF for  $\mathbf{y} = (0.45, 0.45, 0.4207)$  m.

beyond  $A'$  will be erroneous due to the  $1/j_a(\cdot)$  term in (17). In order to minimize these errors, the higher order modes can be discarded as they are simply absent in the actual soundfield. The same technique can be applied to the larger microphone array when calculating the soundfield at  $\eta$ . When the inactive modes are discarded as explained above, the matrix dimensions of (24) will vary with varying frequency. An extensive study on this solution and the resulting improvement in array robustness to noise is presented in [38] for the 2D case.

In order to verify the geometrical flexibility of the proposed parameterization, we repeated the same process for a different configuration of  $\eta$  and  $\zeta$ . This was done by re-defining the vector from  $\mathbf{O}$  to  $\mathbf{O}_s$  as  $\mathbf{R}_{sr} = (0.3 \ 0.3 \ 0.3)$  m which resulted in  $\eta$  and  $\zeta$  to overlap on each other. However, all other design parameters were remained the same, which added no changes to the loudspeaker and microphone array parameters. Figures 6(a) and 6(b) show the actual and reconstructed RTF between a source at  $\mathbf{y} = (0.35, 0.35, 0.3707)$  m and all points in the receiver region for a frequency of  $f = 900$  Hz (the circle represents the receiver region). Similarly, Figs. 6(c) and 6(d) shows the corresponding results for a secondary source at  $\mathbf{y} = (0.45, 0.45, 0.4207)$  m. The reconstructed results in both cases appear almost error-less verifying the geometrical flexibility of the proposed parameterization. However, when  $\eta$  and  $\zeta$  overlap on each other, extra caution should be taken to avoid potential conflicts between the loudspeaker and microphone locations. Furthermore, when a loudspeaker is too near to a microphone, there will be potential errors stemmed from nearfield truncation [39].

<sup>8</sup>One-to-one combinations meaning, the first source location paired with the first receiver location, the second source location paired with the second receiver location etc.

## V. CONCLUSION

We have introduced a novel method to parameterize the three dimensional RTF between two arbitrary points from a sizeable spatial region where the source(s) lie and a sizeable spatial region where the receiver(s) lie. The modal based parameterization only requires a finite number of RTF coefficients to describe an infinite number of RTFs between the two regions and therefore, it can also be used to characterize an entire room at once. However, when an arbitrary shaped room is being measured for RTF parameterization, the corresponding microphone and loudspeaker array geometries may be altered to a similar geometry along (or close to) the room walls. We also presented a practical method of extracting the RTF coefficients which only requires a single loudspeaker and a single microphone, provided the room acoustics remain stationary. This result substantially simplifies the problem of room equalization by simplifying the RTF measurements. It also has the added advantage of providing robustness to both source variations as well as receiver variations.

## REFERENCES

- [1] H. Kuttruff, *Room acoustics*. CRC Press, 2009.
- [2] J. Mourjopoulos and M. Paraskevas, "Pole and zero modeling of room transfer functions," *Journal of Sound and Vibration*, vol. 146, no. 2, pp. 281–302, 1991.
- [3] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *Speech and Audio Processing, IEEE Transactions on*, vol. 2, no. 2, pp. 320–328, 1994.
- [4] B. Radlovic, R. Williamson, and R. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, 2000.
- [5] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.
- [6] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, 1979.
- [7] R. Duraiswami, D. Zotkin, and N. Gumerov, "Fast evaluation of the room transfer function using multipole expansion," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 565–576, 2007.
- [8] A. Krokstad, S. Strom, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *Journal of Sound and Vibration*, vol. 8, no. 1, pp. 118–125, 1968.
- [9] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan, "A beam tracing method for interactive architectural acoustics," *The Journal of the acoustical society of America*, vol. 115, no. 2, pp. 739–756, 2004.
- [10] M. Hodgson and E. Nosal, "Experimental evaluation of radiosity for room sound-field prediction," *The Journal of the Acoustical Society of America*, vol. 120, no. 2, pp. 808–819, 2006.
- [11] D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems," *The Journal of the Acoustical Society of America*, vol. 98, no. 6, pp. 3302–3308, 1995.
- [12] K. Kowalczyk and M. Wästjén, "Room acoustics simulation using 3-d compact explicit ftdt schemes," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 34–46, 2011.
- [13] S. Siltanen, T. Lokki, S. Kiminki, and L. Savioja, "The room acoustic rendering equation," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1624–1635, 2007.
- [14] A. Southern, S. Siltanen, and L. Savioja, "Spatial room impulse responses with a hybrid modeling method," in *Audio Engineering Society Convention 130*. Audio Engineering Society, 2011.
- [15] E. Lehmann and A. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1429–1439, 2010.
- [16] J. Mourjopoulos, "Digital equalization of room acoustics," *Journal of the Audio Engineering Society*, vol. 42, no. 11, pp. 884–900, 1994.
- [17] S. Bharitkar and C. Kyriakakis, "A cluster centroid method for room response equalization at multiple locations," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*. IEEE, 2001, pp. 55–58.
- [18] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 17–28, 2013.
- [19] B. Radlovic, R. Williamson, and R. Kennedy, "On the poor robustness of sound equalization in reverberant environments," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2. IEEE, 1999, pp. 881–884.
- [20] A. Farina, P. Martignon, A. Capra, and S. Fontana, "Measuring impulse responses containing complete spatial information," in *22nd AES-UK Conference*, 2007, pp. 11–12.
- [21] M. Pollow, J. Klein, P. Dietrich, and M. Vorländer, "Including directivity patterns in room acoustical measurements," in *Proceedings of Meetings on Acoustics*, vol. 19, no. 1. Acoustical Society of America, 2013, p. 015008.
- [22] S. Pelzer, M. Pollow, and M. Vorländer, "Continuous and exchangeable directivity patterns in room acoustic simulation."
- [23] A. Farina, "Room impulse responses as temporal and spatial filters," in *The 9th Western Pacific Acoustics Conference, Seoul, Korea*, 2006.
- [24] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, 2005.
- [25] M. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004–1025, 2005.
- [26] P. Samarasinghe, T. Abhayapala, and M. Poletti, "3d spatial soundfield recording over large regions," in *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on*. IEEE, 2012, pp. 1–4.
- [27] M. Schroeder, "The schroeder frequency revisited," *The Journal of the Acoustical Society of America*, vol. 99, no. 5, pp. 3240–3241, 1996.
- [28] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.
- [29] T. Abhayapala and D. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Acoustics, Speech, and Signal Processing (ICASSP), IEEE International Conference on*, vol. II, 2002, pp. 1949–1952.
- [30] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Acoustics, Speech, and Signal Processing (ICASSP), IEEE International Conference on*, vol. 2, 2002, pp. 1781–1784.
- [31] B. Rafaely, "The spherical shell microphone array," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 16, pp. 740–747, 2008.
- [32] I. Balmages and B. Rafaely, "Open-sphere designs for spherical microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 727–732, 2007.
- [33] F. M. Fazi and P. A. Nelson, "Nonuniqueness of the solution of the sound field reproduction problem with boundary pressure control," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 1–14, 2012.
- [34] M. Poletti, T. Abhayapala, and P. Samarasinghe, "Interior and exterior sound field control using two dimensional higher-order variable-directivity sources," *The Journal of the Acoustical Society of America*, vol. 131, no. 5, pp. 3814–3823, 2012.
- [35] P. Samarasinghe, M. Poletti, S. Salehin, T. Abhayapala, and F. Fazi, "3d soundfield reproduction using higher order loudspeakers," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2013, pp. 306–310.
- [36] Z. Li, R. Duraiswami, and N. Gumerov, "Capture and recreation of higher order 3d sound fields via reciprocity," in *International Conference on Audio Display (ICAD)*, 2004, pp. 6–9.
- [37] J. Fiege, "Integration nodes for the sphere," <http://www.mathematik.uni-dortmund.de/lx/research/projects/fliege/nodes/nodes.html>, 2009.
- [38] P. Samarasinghe, T. Abhayapala, and M. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 3, pp. 647–658, 2014.
- [39] S. Koc, J. Song, and W. C. Chew, "Error analysis for the numerical evaluation of the diagonal forms of the scalar spherical addition theorem," *SIAM journal on numerical analysis*, vol. 36, no. 3, pp. 906–921, 1999.